Title: Partner Selection into Policy Relevant Field Experiments

Author: Michèle Belot Jonathan James

PII: S0167-2681(15)00331-5
DOI: http://dx.doi.org/doi:10.1016/j.jebo.2015.12.007
Reference: JEBO 3728

To appear in: *Journal of Economic Behavior & Organization*

Received date: 7-4-2015
Revised date: 17-12-2015
Accepted date: 18-12-2015

Partner Selection into Policy Relevant Field Experiments*

Michèle Belot (University of Edinburgh).
Jonathan James (University of Bath)

April 7, 2015

JEL Classification: C93, I18, J13

Keywords: Selection, Field Experiments, Randomised controlled trials, External Validity

*Corresponding author: Jonathan James, email: j.james@bath.ac.uk, tel: 07786260013

- We study investigates the issue of self-selection of stakeholders into participation and collaboration in policy-relevant experiments.
- We document and test the implications of self-selection in the context of randomised policy experiment we conducted in primary schools in the UK.
- The experiment tests the effects of an intervention aimed at encouraging children to make more healthy choices at lunch.
- There is mild evidence of selection on key observables such as obesity levels and socio-economic characteristics.
- There is selection along indicators of involvement in healthy lifestyle programmes at the school level, but the magnitude is small.

Partner Selection into Policy Relevant Field Experiments*

Michèle Belot (University of Edinburgh).

Jonathan James (University of Bath)

December 17, 2015

Abstract

This study investigates the issue of self-selection of stakeholders into participation and collaboration in policy-relevant experiments. We document and test the implications of self-selection in the context of randomised policy experiment we conducted in primary schools in the UK. The main questions we ask are (1) is there evidence of selection on key observable characteristics likely to matter for the outcome of interest and (2) to what extent does selection matter. The experimental work consists in testing the effects of an intervention aimed at encouraging children to make more healthy choices at lunch. We recruited schools through local authorities and randomised schools across two incentive treatments and a control group. We document the selection taking place both at the level of local authorities and at the school level. Overall we find mild evidence of selection on key observables such as obesity levels and socio-economic characteristics. We find evidence of selection along indicators of involvement in healthy lifestyle programmes at the school level, but the magnitude is small.

JEL Classification: C93, I18, J13

Keywords: Selection, Field Experiments, Randomised controlled trials, External Validity

*We thank the Esmee Fairbairn Foundation for their financial support (grant number 10-2206). We thank seminar participants at the University of Bath, Oxford, Stirling, and conference participants at the European Economic Association conference (Malaga, 2012), Royal Economic Society conference (Royal Holloway, London, 2013), Florence Workshop on Behavioural and Experimental Economics (May 2013), International Conference on Public Policy (Grenoble, June 2013).

1. Introduction

Field experiments in economics and the social sciences have become increasing popular (Holt, 2005). The main driving factors behind this increasing prevalence are, on the one hand, the quest for identification of causal mechanisms – which is easier to achieve when researchers are directly involved in manipulating the economic environment of interest – and, on the other hand, a quest to remain close to reality as opposed to studying subjects in an isolated laboratory context. There is now a stronghold of researchers advocating the case for randomized controlled trials (RCTs) in social policy (Burtless, 1995, Duflo and Kremer, 2005).

This study is interested specifically in a fundamental methodological issue associated with field experimental research: the selection of field collaborators into the experiment. Conducting field experiments usually requires finding collaborators such as employers, policymakers, schools, etc. who are prepared to collaborate with researchers and provide the necessary support for data collection. As List (2011) puts it, the support of a key person prepared to stand behind the research project is often critical: "*Have a champion within the organization – the higher up the better. Making the experiment a "we" project instead of an "us versus them" pursuit as early as possible is critical.*" Surprisingly perhaps, field experimental papers devote very little attention to the issue of "selection into the experiment". As an illustration, we provide in Table A1 a brief overview of the information provided in field experimental studies published in the top 5 journals and in the American Economic Journal: Applied Economics. We focus on the fields of policy evaluation, personnel economics and development economics, which have all experienced a significant increase in the popularity of field experimental research.[1] In most cases we know little or nothing about how the collaborators were selected and approached, and the experimental sample is not compared to the broader population of interest. One notable exception is a recent paper by Fryer (2011).

Researchers are well aware of the limitations that result from restricting experiments to subjects who have opted in. The main limitation is a possible selection bias and a lack of external validity. Of course, some form of selection is inevitable. An experiment will, for example, take place within a given geographical area and at a particular point in time. This initial selection is often for practical reasons. Researchers located in California will find it more practical to conduct a field experiment in California than across the entire United States or across several countries. It is probably even desirable that the experiment can be conducted with sufficient oversight by the researchers. It does nevertheless raise

[1] We searched these journals systematically for the keywords 'field experiment'.

issues of generalizability, insofar as it restricts the sample to a population with certain characteristics (e.g. people living in California at a particular time).

How important is selection in field experimental research? Harrison and List (2004) acknowledge in their review paper that we know very little about the implications of self-selection for field experimental research. Seven years later Ludwig et al. (2011) point out that this is still an open question that has not been answered. At the moment, researchers tend to be very conservative and modest in their claims regarding the external validity of their results. We believe that a proper documentation of the selection process would help us drawing more general lessons from field randomised controlled experiments.

We study self-selection in participation in an experiment conducted in a highly policy relevant domain: children's diet. The goal of the experiment is to test the effectiveness of various incentive schemes to encourage children to eat fruit and vegetables at lunch[2]. To conduct this experiment, we sought the collaboration of primary schools in England. We recruited them through local education authorities (LEAs), which play an overarching and coordinating role. Rather than picking a set of local authorities in an arbitrary manner, we approached all local authorities in the country at the same time and in the same manner (via e-mail) and asked whether they would be interested in collaborating with us. If they responded positively, we asked them to bring us in contact with at least five local schools representative of the local authority. Providing names of schools requires some effort, so the type of selection we study is not only based on initial interest (that is low cost to indicate) but also on actual commitment in the experiment. The randomisation eventually takes place at the school level and within local authorities, so local authorities should expect some schools to be treated and some schools to be part of a control group. The schools are the ones that are ultimately directly involved in the experiment and data collection. We contacted the schools suggested by the local authorities, briefed them about the project and they then decided whether to participate or not. Thus, we have potential selection operating at different levels: self-selection of local authorities, selection of "representative" schools by local authorities and selection of schools into the experiment.

We document how selection operates at these different levels along observable characteristics of the population under consideration – characteristics that we would expect could matter in the decision to participate, such as obesity rates and socio-economic indicators. We consider a wide range of variables that could a priori be relevant and see whether they are correlated with selection or not. Then we investigate whether the

[2]We refer to Belot et al. (2015) for the full analysis of the experiment.

treatment effects are biased by observables correlated with selection, we do not find any evidence that this is the case.

It is clear that we cannot control for some key variables (such as personality characteristics of the people involved) that may introduce a selection bias in the RCT as well. Moreover, it is clear that if selection was only driven by observables, it would be straightforward to correct for it. The point here is to get a sense of how much selection takes place along characteristics that could a priori be relevant and are observable, which should in principle give a sense of the importance of the full selection problem.

Our findings can be summarised as follows. First, we find that out of the 150 local authorities we initially contacted, only 12 eventually participated in the actual experiment. We find that local authorities who express initial interest tend to be larger and richer, and have less favourable characteristics in terms of the outcome of interest (e.g. lower rates of fruit and vegetable consumption), but the selection is mild. We find no evidence that selection operates according to pre-trends. We also find little evidence of selection at the school level, except for one dimension, which is that the schools suggested by the local authorities to conduct the experiment are more likely to be involved in programmes promoting healthy lifestyles (as evaluated by an independent official body). Second, we do not find any significant correlations between the treatment effects of the experiment and the variables which, albeit to a mild degree, are correlated with selection into the experiment.

Of course, the selection we document here is also "case-specific". We cannot claim that the selection we document informs us about the magnitude and type of selection taking place in other policy-relevant field experiments. But in the domain of policy-relevant field experiments, this is an example of an intervention that targets a"hot" topic on the policy agenda (children's obesity) and the experimental intervention we propose is typical in the sense that we propose to compensate the partners for the costs involved with conducting the intervention and collecting the data. Even in that case we find that only few potential partners eventually participate in the experiment. We believe this is typical of many policy-relevant field experiments. If only a selected few eventually participate in the study, we owe to pay more attention to how this selection affects our results.

We propose that the documentation exercise we perform here should be a minimum standard for field experimental research in social sciences. It seems obvious that collecting information on key variables at the early stages of an experiment can provide useful insights into the possible magnitude of selection and is helpful to see how the results can be generalized to the population at large, which is the overriding aim of pilot studies and policy-relevant experiments.

The remain of the paper is structured as follows. Section 2 discusses related work. Section 3 presents a simple model of the problem we are examining. Section 4 describes the experimental design and the context of our study. In Section 5, we look at the determinants of selection at the local authority and school levels. In Section 6 we examine how the treatment effects from the experiment varying according to the determinants of selection. We conclude in Section 7.

2. Related work

There is surprisingly little systematic work studying selection into field experiments. The work most related to ours is the work on selection issues arising in policy evaluation studies (Heckman and Vytlacil (2006)). This work derives theoretical implications regarding the sign of the selection bias one should expect in these interventions. The general prediction is that one should expect a positive selection bias: those who self-select in the intervention are likely to be those who expect to benefit most from the intervention.

Malani (2008) develops a model of self-selection into medical randomised controlled trials where participants self-select into the experiment based on a prior belief about the effectiveness of the treatment. He assumes that patients have a choice between an old treatment and participating in a randomised controlled trial (RCT) involving a new treatment. As the probability of being assigned to the treatment group in the RCT increases, less optimistic patients, those with a lower expected treatment effect, are willing to participate in the RCT. If the probability is lower than 1 (which will always be the case by definition in a RCT), only the most optimistic will be willing to experiment and we will obtain a positive selection bias. The main difference with our model is that people cannot access the new treatment out of the RCT. While this may be the case in medical trials, it is usually not the case for experiments conducted in the social sciences. This is why those who are optimistic may opt out and prefer to implement the intervention (or a similar intervention) themselves.

In a related paper (Belot and James, 2014), we extend Malani's model to the context of randomised social and policy experiments. A key difference is that in a social or policy experiment, the decision to participate in a randomized controlled trial does not only entail a choice between participating or not doing anything. A third alternative could be to conduct a similar intervention without taking part and without running the risk of being in the control group. This alternative is likely to be attractive for the most optimistic - those who believe the intervention may be effective. They may prefer to opt out from the experiment because there is a chance they will end up in the control group and not receive the benefits from the intervention. In our context, the idea of using incentives in a health-related context is not new. In fact there is a growing market for stickers and other rewards to encourage children to eat healthily. Thus, it is plausible that schools would have wanted to implement an incentive scheme independently of being approached by us. But of course if they participate and end up in the control group, we will explicitly ask them not to implement any incentive scheme over the course of the study. This example is, we believe, quite representative for many randomised controlled experiments. Thus, two types of selection (positive and negative) could take place at the same time. On the one hand, we have a group of "optimists" (who expect the treatment to be successful) who could opt out and introduce a negative selection bias. On the other hand, we have a group of "pessimists" (who do not expect the treatment to be successful), who could opt out as well if they are not fully compensated for the experimentation and implementation costs. Their decision not to participate will introduce a positive selection bias. Because both types of selection could take place at the same time, it is not possible to sign the direction of the selection bias.

Allcott (2015) examines the issue of site selection bias and external validity. He uses the Opower energy conservation programme which provided Home Energy Reports comparing energy use with their neighbours and well as providing tips on conversing energy. He uses the first 10 replications of the programme which were spread across the country covering over half a million households to predict the first-year effect in the forthcoming 101 sites. Using the microdata from these 10 initial sites over predicts the mean average treatment effect for the remaining sites with the early sites being positively selected.

He also discusses various mechanisms of partner selection bias in randomized controlled trials (RCTs) – both positive and negative. There could be a positive bias because implementing randomized trials requires managerial ability and operational efficacy and partners who run the most effective programmes may also be the best equipped to evaluate these programme. This form of positive partner selection bias has been called "gold plating" (Duflo, Glennerster, and Kremer 2007). Another form of positive partner selection bias results from the fact that those running effective programmes are more keen to show they work than those who fear they are running ineffective ones, and prefer to remain "ignorant" (Pritchett 2002). Negative partner selection bias could arise if those who are running the most effective treatments have already treated the parts of their population that have the largest treatment effects ("diminishing returns bias").

More generally, a few recent studies study the issue of selection in various types of experimental research. Gautier and Van Der Klaauw (2012) examine selection of individual participants in a field experiment setting. They examine the effect of selection in a gift exchange game by exploiting a promotion in all hotels of a chain in Belgium and the Netherlands that allowed customers to pay what they wished for a one-night stay. They compare the behaviour of two groups of participants, an involuntary group who had booked before the announcement of the promotion and were not aware of it at the time of booking with a group who booked after the promotion was announced - a voluntary group. The self-selected voluntary group paid less and did not respond to an exogenous change in the posted price unlike the involuntary group. These results suggest that experiments aiming at testing for pro-social behaviour are underestimating its presence in particular where participants can self-select into the experiment. Although this does not involve partner selection it shows in a unique and natural setting the role that self-selection can play in leading to biased results in field (and lab) experiments. Frijters et al. (2015) compare participants to an artefactual field experiment in urban China with the survey population of migrants from which they were recruited. They find that the experimental participants differ significantly in their education, in the extent to which they lend money to friends and the amount of hours worked. Cleave et al. (2013) study selection into lab experiments by comparing the behaviour of students in a classroom experiment to students participating to a "standard" laboratory experiment. Both samples play a trust game and a lottery choice task. They do not significant differences across these two samples. However, they do find that people who sent less in a trust game were more likely to participate in a laboratory experiment. Similarly, Slonim et al. (2013) finds lab participants have lower income, work fewer hours, and volunteer more often. Overall they find that people who come to the lab display behaviours that are correlated with interest in experiments. To summarise, there are few studies documenting the issue of self selection in randomised controlled experiments in social sciences and, in theory, the sign of the selection bias could go either way.

3. Model and Empirical Strategy

3.1 The issue of external validity

We first discuss the standard assumptions required for external validity, following Hotz, Imbens, and Mortimer (2005) and Allcott (2015). Assuming a population of N units, with i indicating an individual unit (such as a school). $T_i \in \{0, 1\}$ is a dummy variable indicating whether the unit is treated ($T_i = 1$) or not ($T_i = 0$). Each individual unit has two potential outcomes, $Y_i(1)$ if treated and $Y_i(0)$ if not treated. X_i is a vector of observable pre-treatment covariates, such as school test scores, prevalence of obesity, etc. We will come back to these possible determinants later on in the section. The treatment effect is defined as $\tau_i = Y_i(1) - Y_i(0)$.The heterogeneity in treatment effects could be driven by factors such as the characteristics of the pupils, the motivation of the school staff or the experience with similar interventions.

Suppose D_i is a dummy variable indicating whether the unit participates to the randomized controlled trial ($D_i = 1$) or not ($D_i = 0$). We are interested in the average treatment effect (ATE). The ATE can be consistently estimated under the following assumptions:

(1) Unconfoundedness. $T_i \perp (Y_i(1), Y_i(0))|X_i$. That is, the treatment must have been assigned at random, conditionally on the vector of observable pre-treatment characteristics.

(2) Overlap. $0 < Pr(T_i = 1|X_i = x) < 1$. *Thus, no covariate value, or combination of covariate values, perfectly predicts getting the treatment.*

(3) External unconfoundedness of participation. $D_i \perp (Y_i(1), Y_i(0))|X_i$.That is, participation in the experiment must have also been randomised, conditionally on pre-treatment variables X_i.

(4) External Overlap. $0 < Pr(D_i = 1|X_i = x) < 1$. This states that no covariate value or combinaion of covariate values perfectly predicts participation in the experiment.

Allcott (2015) discusses an additional assumption relevant to a situation where many units participate in the RCT and researchers wish to extrapolate from these many units to a large set of target units.

(5) No unit selection bias over a large number of units. $E(\tau_i(X)|D_i = 1) = E(\tau_i(X)|D_i = 0)$. This assumption is useful because it allows heterogeneity in τ_i across units as long as they average out between participating and non-participating units. He is able to test this assumption directly.

3.2 A simple model of self-selection

We will now propose a simple theoretical framework to illustrate the type of selection that may take place in the context of this experiment. The goal of the model is to highlight a mechanism for self-selection that has not received much attention so far in the literature. Specifically, we will show that positive and negative selection can arise at the same time.

We refer to Belot and James (2014) for a more general presentation of the model. Here we present how the same model applies to this particular setting.

We assume schools decide whether to participate or not to RCT. Schools do not observe τ_i ex ante, but we assume they have a prior belief about the effectiveness of the treatment, which we denote by τ_i^*. We assume that these beliefs are on average unbiased, i.e. $E(\tau_i^*) = \tau_i$. Variance in beliefs (for a given τ_i) could be driven by chance, but also by other factors (most likely unobservable to the econometrician) such as over- or under-confidence of the decision-maker or trust in the researchers who propose the experiment.

In the context of our experiment, conducting the intervention and participating to the experiment entail costs, even for the control group. First, there are costs that are incurred by all units participating, whether there are in the treatment or control groups. These costs are associated with meetings with researchers, data collection, logistic organisation, etc, which would not be incurred if they did not participate. We denote this cost c_E. Second, there is the cost of conducting the intervention, which is associated with the implementation of the intervention itself and is only incurred by the treatment group. We denote this cost c_I. We assume that if units would incur c_I as well if they would carry out the intervention on their own.

We offered a compensation both to conduct the intervention (i.e. we paid for all the materials involved) and for data collection (we paid for monitoring and for postage). However, it is possible that schools were not fully compensated for their time and efforts. We denote these subsidies s_E (subsidies to experimental costs) and s_I (subsidies to the intervention).

Schools are assigned to the treatment group with probably π (in our case, it was 2/3) and to the control group with probability $(1-\pi)$. The treatment group will implement the intervention, while the control group is told not to implement any competing or similar intervention. On the other hand, the non participants could of course implement alternative interventions (similar or not to the experimental intervention). We assume that schools know π (this was the case in our setting).[3]

Thus, for participants the expected benefit from participating is equal to $\pi\ (\tau_i^* + s_I - c_I) + s_E - c_E$. For non participants, the payoff is equal to $max(\tau_i^* - c_I, 0)$.

The most interesting case is one where the best (i.e. most cost effective) available intervention outside the experiment is the experimental intervention itself (or an inter-

[3]Note that if it was not known to decision makers, we would not expect changes in the results as long as beliefs are unbiased. In the absence of information, decision markers are, for example, likely to assume that $\pi_i = 0.5$, in which case their beliefs could be systematically biased and biases in beliefs could affect the decision to participate as well.

vention that is similar to it in the sense that there is a positive correlation in treatment effects). In that case, the two conditions for participating to the experiment are as follows:

$$\begin{aligned} \pi(\tau_i^* + s_I - c_I) + s_E - c_E &> \tau_i^* - c_I \\ &\Longleftrightarrow \\ \tau_i^* &< \frac{s_E + \pi s_I - c_E}{(1-\pi)} - c_I \end{aligned} \tag{1}$$

Condition (1) means that participating must be more profitable in expectations than conducting the intervention outside the experiment. It shows that the most optimistic opt out because they prefer to conduct the intervention outside the experiment. This type of selection decreases with π, the probability of being assigned to the treatment group and with the subsidies (s_E and s_I), but increases with the costs of participating to the experiment (c_E) and of conducting the intervention (c_I). The novelty here is to show that *negative* selection can take place.

At the same time, the more "standard" positive selection, commonly discussed in policy evaluation studies (Heckman and Vytlacil (2006)) could take place as well. Partners who have pessimistic beliefs (such that $\tau_i^* < c_I$), will opt in only if

$$\begin{aligned} \pi(\tau_i^* + s_I - c_I) + s_E - c_E &> 0 \\ &\Longleftrightarrow \\ \tau_i^* &> c_E - s_E + \pi(c_I - s_I) \end{aligned} \tag{2}$$

Condition (2) is more likely to be satistied if the subsidies are high (s_E and s_I) and if the costs of participating to the experiment (c_E) and the costs of conducting the intervention (c_I) are low.

Note that in our setting, we contacted local authorities that represent schools. Thus, we should in fact talk about self selection by local authorities. The problem is isomorphic to this one insofar as we assume that local authorities represent the average school. That is, they take into account that a proportion π of their schools will be assigned to the treatment group and the remaining will be assigned to the control group.

Summarising, the model predicts that both negative and positive selection can take place at the same time and that many variables that are usually thought of increasing positive selection could actually lead to negative selection as well. Specifically, higher costs will lead to more positive and more negative selection.

Note that the model only allows for heterogeneity in the effectiveness of treatment. It would be straightforward to extend it though to allow for heterogeneity in costs. The surprising result here is that, again, higher costs increase both types of selection (positive and negative) at the same time: If the costs are relatively high (holding the subsidies constant), the benefit of participating to the experiment, in comparison to conducting the intervention outside the experiment, decreases. But higher costs will also discourage units to conduct the intervention at all.

Extensions. The model could be extended further to take account of a richer objective function for the potential experimental units. As it is, we effectively assume that units care about the effectiveness of the intervention and do not differ in the way they care about it. In practice and in our context in particular, we could expect that some schools may care more about the outcome (i.e. getting children to eat more fruit and vegetables) than others. Schools could also be more or less risk averse. That is, we could have a utility function $u_i(\tau_i^*)$ that captures the degree to which they care and their risk aversion. The utility function could include altruism and risk aversion parameters. The implications of this extension are also relatively simple: Any parameter that increases the utility of the school for a given τ_i^* could lead to both positive and negative selection. For example, principals who care a lot about the outcome may be more likely to opt out if they are optimistic about the success of the intervention because conducting it outside the experiment increases the chances of benefitting from it. On the other hand, if they are relatively pessimistic about the effectiveness (and would therefore not conduct the intervention on their own), the more they care about the outcome, the more likely they are to opt out of the experiment. To our knowledge, such predictions are novel and highlight that the type of selection taking place is not straightforward to predict. In contrast to predictions based on models in the line of Heckman and Vytlacil (2006), participating to the experiment does not necessarily draw the most motivated or altruistic partners. And the main reason for that is that participating to the experiment entails an important opportunity cost (having a chance of being in the control group), which has received little attention so far.

Ultimately, one can only rely on careful empirical documentation of the selection taking place to conjecture on which type of selection (positive or negative) may be dominating.

3.3 Empirical strategy

As discussed above, the main determinants of selection included in the model are beliefs about the effectiveness of the treatment, probabilities of being assigned to treatment and control groups, costs of carrying out the intervention and of participating to the experiment and the subsidies we will provide. Of course, the question is how do we know whether there is positive or negative selection? By definition, we do not observe the treatment effects for individuals who chose not to participate in the experiment. One strategy is to exploit the information on observables at the time of selection and pose assumptions regarding the relationship between observables and the variables determining selection.

In the context of our study, we think of three main categories of (observable) factors that could affect the marginal benefit from intervening and the costs associated with it and with the experiment:

- Characteristics of the population targeted, such as the pre-intervention levels of fruit and vegetable consumption. Suppose the outcome is a normal good with decreasing marginal utility, then all else equal we would expect those with lower levels of fruit and vegetable consumption to have higher τ_i. We have information on consumption of fruit and vegetables pre-intervention, as well as levels of obesity across local authorities. We also include a set of variables measuring average pupil performance (standardised test scores), which are likely to be correlated with health-related behaviours.

- The degree to which individuals are able to implement the intervention, which will vary the costs of implementing the intervention. Here we think of organisatory and planning skills that may be required to implement the intervention. If the potential partners are larger in size and have access to more resources, they may be more effective in implementing the intervention and in collaborating with the researchers. We have information on size of local authorities and schools, as well as indicators of income per capita at the local authority level and percentage of children receiving free school meals. We would expect richer and larger authorities to have more capacity (i.e. we conjecture there are increasing returns to scale) and may therefore have a higher τ_i.

- The degree to which individuals care about the outcome of interest. We have information on the degree to which schools promote healthy lifestyles. Also, one would expect that areas with higher obesity rates may be more concerned with implementing policies aimed at encouraging healthier lifestyles.

We now turn to the description of the context and experimental design.

4. Context and Experimental Design

4.1 Policy Context

Obesity and diet are high on the policy agenda. The broad question of our study (which we used to approach the local public authorities), is whether providing short-term incentives can successfully encourage children to make healthier nutritional choices and develop healthier habits. We contacted all 150 local authorities in the United Kingdom on Friday 2 July 2010 and Monday 5 July 2010 (we randomly selected half of the sample to be sent on each day). The e-mail (see appendix D) described who we are and the aim of the project. We specifically indicated that we were interested in comparing the effectiveness of incentive schemes to increase the consumption of fruit and vegetables at lunch in schools and that the interventions were designed to target children who have been found to respond little to health interventions, such as children from poorer socio-economic backgrounds and boys. These specific observables (consumption of fruit and vegetables, children from poorer socio-economic backgrounds) vary across local authorities and we are able to document precisely how selection takes place based on the observables we mentioned in the letter.

The timing of this initial contact coincided with the First National Child Obesity Week[4] taking place in the UK, from July 5th until July 10th. School meals and children's diet were also widely discussed in the media. On 30 June 2010, the then Secretary of Health Andrew Lansley criticized the Jamie Oliver School Food revolution, referring to the campaign led by the British chef Jamie Oliver aimed at improving parental cooking skills across the country (the criticism was targeted at the patronizing approach of the campaign). The public criticism triggered a strong reaction in the media and the public. Lansley apologised to Oliver later in October 2010. At the same time, the climate regarding public funding was relatively grim. Large public funding cuts were expected across the whole country.

[4]http://www.mendcentral.org/ncow

4.2 Experimental Design

The experiment is fully documented in Belot, James and Nolen (2015). We will only briefly describe the key elements of the experiment here. The goal of the experiment is to study the effects of providing incentives on choosing and eating healthy items using two different interventions: an individual scheme and competition scheme. In both schemes, pupils were given a sticker for choosing or bringing in a healthy item at lunch. Then, at the end of the week (Friday afternoon), each student had the opportunity to win a larger prize. In the individual scheme, pupils were eligible for an additional reward if they collected four stickers or more during the week. In the competition scheme, students were assigned to random groups of four, and the pupil with the most stickers in each group won an additional reward.

Our initial e-mail mentioned that we intended to test the effectiveness of incentive schemes, without being specific. We also mentioned being particularly interested in studying how boys would respond in comparison to girls, and how pupils from disadvantaged backgrounds would respond in comparison to pupils from richer backgrounds. These specific interests came as a follow-up from a previous study we conducted in the UK (Belot and James, 2011).

We now turn to the documentation of the selection in the context of our study.

5. Determinants of Selection

5.1 Selection at the level of local authorities

We first study the determinants of selection at the local authority level. Figure 1 presents the decision process showing the different types of selection. We differentiate between four different stages of selection. First, the local authorities responded to our e-mail or not. Second, their response may have been positive ("positive initial interest"). Third, we examine whether the local authority collaborated with us. We asked each local authority to provide the names of 5 schools that could take part in the experiment; and asked for these schools to be as representative as possible of the schools in the local area, in terms of socio-demographics. The LEAs did not all respond to this request in the same way. Some did contact schools before providing names; others recommended to advertise in all schools of their area (and helped us doing that) and others provided a list without contacting them first. We classify all these responses as collaboration since they did take action to address our request. The final determinant of selection is participation. We define a local authority as participating as if they have schools taking part in the intervention. Out of the 150 local authorities contacted, 63 responded within a month. 33 indicated a positive interest in collaborating and 18 asked for a meeting. 22 of them have effectively engaged in the first step of collaboration. 12 LEAs finally participated in the actual experiment.

5.2 Contemporary indicators

We first consider socio-economic and health indicators which are most associated with the proposed experiment. We have these at the local authority level for 2008, which were the most recent available indicators in June 2010 when the recruitment of the experiment took place. The next section discusses the role of pre-intervention trends.

We have information on the average weekly household income[5], the percentage of free school meals (school meals are part of a means-tested programme and around 17% of children receive free school meals in the UK), the number of schools in the area, and a number of variables capturing overweight children and obesity rates, as well as adult health habits (and in particular, the percentage of people eating the recommended five portions of fruit and vegetables per day).

Regarding the initial response first (Table 2 Panel A) – a response can be either negative (0) or positive (1). Each of our definitions of selection are estimated using a probit regression, and the estimates in each of the tables are presented as marginal effects. We find that areas with a larger number of schools were more likely to respond, and we find some evidence (although the coefficients are not always significant) that authorities with a higher income and a lower consumption of fruit and vegetables were more likely to do so. Turning to the probability of a positive interested response (Table 2 panel B) (and assuming that no response within a month is a negative response), we find that larger LEAs and LEAs with richer households are more likely to be interested. We find that those with a lower consumption of fruit and vegetables are also more likely to respond, as well as those more confronted with a child obesity problem. These results support the hypothesis that areas with lower levels of fruit and vegetable consumption and higher obesity rates may care more about the outcome of interest, and may have higher marginal benefits from the intervention.

We include additional health indicators to gauge whether the selection is driven by the general state of health in the local area or whether it operates mainly through the indicators and variables that are specifically targeted in the study (children's obesity and diet). Of course these general indicators are correlated with the more specific ones, but the question is whether these general indicators alone are good predictors of interest. In columns

[5]Full description of the variables and their sources are presented in Appendix C.

(5) and (6) we substitute indicators for smoking and binge drinking with the indicator of fruit and vegetable consumption. We find that these indicators are poor predictors of collaboration – the coefficients are close to zero and rather precisely estimated. This suggests that the interest is not reflecting a general poor state of health in the area but is rather reflecting specific weaknesses with respect to the indicators specifically targeted in the study.

Examining those who collaborate with us, Table 3 panel A, we find less systematic differences in terms of observable characteristics than for the initial interest. The number of schools, and income remain significant. Obesity rates are not systematically different. We also fail to find systematic differences according to the percentage of free school meal children. There remains a systematic difference, albeit to a lesser extent, in terms of fruit and vegetable consumption, which is perhaps the most obvious outcome of interest in our study. Table 3 Panel B shows little selection along the lines of actual participation.

Tables B1-B4 examine determinants that, although not directly associated, could determine response and interest in the intervention. We include the level of education at the age groups we would target in the intervention – key stage 1 (taken at aged 6) and key stage 2 (taken at aged 11). Those with better key stage 2 scores are more likely to respond, however, this is not the case for expressing a positive interest. Given the political timing in which we contacted the local authorities, there was a huge amount of uncertainty surrounding funding. We therefore consider three types of public spending measures: Per pupil spending for the current year, the change in spending, and the spending at the local authority level, none of which determine participation at any level. We also control for the gender of the initial two contacts (the CEO of the local authority and the director of children's services), and of the most likely contact to whom our original letter would have been passed on to, the healthy schools contact. The gender of these contacts does not determine response, interest or collaboration.

Summarizing the results, we find some evidence of selection taking place, but it is mild, certainly when it comes to the actual collaboration.

5.3 Pre-trends

We now consider the role of pre-trends in health and education indicators. The question is whether the areas that responded and engaged in collaboration with us were on different trends than others. For example, are these areas confronted with a deterioration in these health indicators or, on the contrary, are they the ones on a positive trend, which could indicate that they are effectively engaged in improving children's diet and obesity already?

Table 4 panel A controls for trends in obesity and overweight rates. We find no evidence of any selection based on those trends. Of course it could be that both types are more likely to self-select: more than one standard deviation up or one than more standard deviation down in comparison to 2006 level. Panel B controls for trend in education, again we find no evidence of self-selection.

5.4 Non monotonicity

As discussed earlier, a priori there could be positive and negative selection taking place at the same time, that is, both the optimists and the pessimists could choose to opt out. To the extent that these beliefs are correlated with observables in a monotonic way, then one way to allow for positive and negative selection at the same time is to allow for non monotonicity in the relationship between observables and the probability of opting in.

Table 5 replaces the continuous measures of the key variables (fruit and vegetable, obesity, and income) with dummies representing the mid-point and an upper value for that variable. These show that those who self-select are those with the greatest problems to solve. The areas with the highest levels of fruit and vegetable consumption are the least likely to express an interest or collaborate. The areas with the top third of fruit and vegetable consumption are between 32 and 29 percentage points more likely to express and interest, around 25 percentage points to collaborate, however there are no systematic differences in the propensity to collaborate or participate.

5.5 Selection of Schools

We next turn to the selection of schools that were recruited into the intervention. As mentioned above we asked each local authority to provide the names of schools that could take part in the experiment. We then contacted these five schools, briefed them and they ultimately decided whether they wanted to participate or not. This allows us to study the selection of schools into the experiment as well as well as examine the variables that might determine non-participation.

We have detailed information about the schools in each local authority. We have information on free school meal participation, school size and a range of school spending variables. In addition, we also have information from school' inspections conducted by the official inspection body Ofsted[6]. These inspections are summarized in reports that provide measures of how well the school operates or performs in various dimensions. There is a general measure of school performance, which assigns scores from 1 to 4, where a score of

[6] www.ofsted.gov.uk/

1 is outstanding, 2 is good, 3 is satisfactory, and 4 is inadequate. There is also a measure that provides information on how well the school performs at *"getting its pupils to adopt healthy lifestyles"*, which is obviously directly relevant in this context. Schools are rated according to a similar scoring rule (from 1 to 4) as the general measure[7]. This measure is the closest indicator we have that captures the involvement of schools in promoting healthy lifestyles.

Table 6 presents the results of the analysis of school selection. To do this, we use the sample of all schools located in the 12 participating local authorities. In columns 1 to 3 the dependent variable in each case is binary represented by a one if the school was chosen and decided to participate in the intervention and zero otherwise. In column 1 we include both infant (for children aged between four and 7) and junior schools.[8]. We find a positive selection with regard to the health promotion measure (Ofsted health score 1) as well as those with a medium level of free school meal participation. However, these are only significant at the 10% level. The second column additionally controls for the average maths and english score at key stage 2[9]. As the test "score" comes from exams which are taken by children aged 11 for key stage 2 column 2 no longer includes those separate infant schools. Column 3 drops the test score variable and shows the estimates only on the sample from column 2. We find a negative selection with regards to an overall indicator of school performance (Ofsted score 1). But the schools that are selected by the local authority are more likely to be those with an "outstanding" score in promoting healthy lifestyles. In column 4 and 5 we instead use the final sample of schools that started and completed the experiment. The pattern is similar for all the selected schools that took part in the experiment.

Next we examine whether there were any differences between schools that participated and schools that initially agreed to participate but dropped out before the start of the

[7]The OFSTED evaluations are not used as an indicator for school headteachers to get a bonus. Nor are the principals are incentivised based on the health behaviours of their pupils, at least not explicitly. However, it has been reported in the media that principals have been feeling under threat of losing their jobs as a result of bad OFSTED reports. See for example: `http://www.theguardian.com/education/2014/mar/11/heads-poor-ofsted-report-dismissal-shortages` Schools (and their governing bodies) have a lot of freedom in determining the principals pay but must follow some rules set by the Department of Education. It is clear that the OFSTED reports are certainly playing a role in evaluating the performance of a school. The evaluation of the principal is part of the OFSTED report (and is separate from the evaluation of other dimensions such as pupil performance and promotion of healthy behaviours).

[8]Infant and junior schools are both types of primary school. Infants schools are for children aged between four and seven and junior schools are for children aged between eight and eleven. Infant are typically linked to a junior school. Some infants schools do however financially report separately

[9]The average point score gives a more rounded picture of how well the students performed. Points are points awarded per subject per pupil along the following lines: working below the level of the test or not awarded 15, level 2 receives 15 points, level 3 gets 21, level 4 gets 27 and 33 points is allocated for level 5. The average score is then calculated from the following: (Total points for English + Total points for maths) / (Total number of eligible pupils for each subject)

experiment. In Table 7 we compare and test the difference in the sample means of the schools that took part in the experiment and those that dropped out and did not start, or complete, the experimental intervention. To determine the significance of the differences, we use the nonparametric Mann-Whitney test. The final column reports the p-value of this test. We do not find any significant differences between the schools that dropped out or did not take part and those which took part in the experiment.

6. Treatment Effects and Observables

In this section we examine the treatment effect of the experiment and examine whether this varies according to the characteristics that have been shown to, albeit mildly, determine selection into the experiment. In particular we focus on the immediate effect of the intervention and focus on how the treatment effect differs according to the proportion of fruit and vegetable consumption for the adult population, the average household income and the number of schools in the local authority. In addition, at the school level we examine the ofsted health score.[10] We estimate the following:

$$\begin{aligned} C_{islt} &= \beta_1 On_t + \beta_2 Comp_s \cdot On_t + \beta_3 Ind_s \cdot On_t \\ &+ \beta_4 Comp_s \cdot On_t \cdot AboveAvgFV_l + \beta_5 Ind_s \cdot On_t \cdot AboveAvgFV_l + \beta_6 On_t \cdot AboveAvgFV_l \\ &+ \beta_7 Comp_s \cdot On_t \cdot HS_s + \beta_8 Ind_s \cdot On_t \cdot HS_s + \beta_8 Ind_s \cdot On \cdot HS_s \\ &+ \beta_{10} Comp_s \cdot On_t \cdot AboveAvgInc_l + \beta_{11} Ind_s \cdot On_t \cdot AboveAvgInc_l + \beta_{12} On_t \cdot AboveAvgInc_l \\ &+ \beta_{13} Comp_s \cdot On_t \cdot AboveAvgNSch_l + \beta_{14} Ind_s \cdot On_t \cdot AboveAvgNSch_l + \beta_{15} On_t \cdot AboveAvgNSch_l \\ &+ \alpha_i + \epsilon_{islt} \end{aligned}$$

where C_{islt} indicates whether individual i in school s located in local authority l on day t chose a fruit or vegetable, we also examine whether the child ate at least some of it (we call this "try"). *On* is a dummy that takes a one for weeks 2 to 5 when the incentives were in place, and 0 for the baseline week (week 1). *Comp* and *Ind* denote whether school s was part of the competition or individual incentive schemes. The treatment effects of interest are given by the interaction terms of *Comp* and *Ind* with *on*. To capture whether the treatment effect differs by the variables that are correlated with selection into the experiment we interact the treatment effects with the additional variables. Specifically, for fruit and vegetable consumption we create a dummy variable, $AboveAvgFV_l$, that

[10] We do not examine the overall ofsted score as we do not have enough schools in each treatment to make this viable.

takes a one if the pupil was in a school in a local authority was above the average for fruit and vegetable consumption compared to other local authorities and interact this with the treatment effect interactions. Furthermore, we include (interacted with the treatment effect) HS_s that takes a one if pupils is in a school that had an ofsted health score that was outstanding and zero otherwise, $AboveAvgInc_l$ taking a one if a pupil lived in a local authority with above average household income, and $AboveAvgNSch_l$ equals one if a pupil lives in a local authority with above average number of schools. The unobserved error term is ϵ_{islt} and an individual fixed effect is given by α_i.

Table 8 presents the estimates from the experiment.[11] The first column presents the estimates for the dependent variable choice[12] for the whole sample. Column 2 considers another group of interest - those who did not choose fruit or vegetable 100% of the time during the baseline week. This group has some margin to improve. In this case we find a larger difference in the treatment effects between those above and below fruit and vegetable consumption. However, when looking at this sub-sample, our analysis contains less than 30 schools. Therefore, the standard clustering methods might not be appropriate. To deal with this we correct for the potential clustering problems using the wild bootstrap method proposed by Cameron, Gelbach, and Miller (2008). The p-values from this procedure are presented in the square brackets. In column 3 we replace the dependent variable with try, and in the final column we keep try as the dependent variable but use the same sample as in column 2 i.e. those who did not choose a fruit or vegetable everyday in the baseline week.

We find that the competition treatment effects are larger for schools with an outstanding ofsted health score, however, this is only statistically significant for the sample that includes those who chose less than 100% in the first week. The treatment effects also significantly differ by average income. Typically the competition treatment effect is negative for areas with above average income but positive for the individual incentive scheme. While none of the variables show a consistent positive treatment effect across both treatments types we do find the joint test of the triple interaction terms to be significant. Therefore, we find some evidence that the treatment effects are significantly different along the lines of the variables that determine selection however it is not always in the same direction.

[11]A full account of the treatment effects of the experiment is documented in Belot, James and Nolen (2015).

[12]The variable is a dummy equal to 1 if the child chose a portion of fruit and vegetables at lunch or brought a fruit or vegetable in their packed lunch.

7. Conclusion

This paper discusses and provides evidence of selection of field experimental partners in a policy relevant experiment. We provide evidence based on a specific policy experiment, which consists of providing incentives to children to eat fruit and vegetables at lunch. All potential partners were approached at the same time. We have access to a range of relevant observable measures likely to be correlated with prior beliefs about the effectiveness of the intervention and with the ability to conduct the intervention. We observe mild selection along observable characteristics. Local authorities interested in collaborating tend to be larger and richer, and have lower levels of fruit and vegetable consumption. Next to that, the schools suggested by local authorities and ultimately participate in the experiment tend to be those who are already outstanding in promoting healthy lifestyles. Local authorities are more likely to involve schools that are already engaged in promoting healthy lifestyles, but who do not score highly on an overall evaluation measure of management. We also do not see a systematic difference between those who were selected by the local authority but do not complete the experiment intervention. Finally, we examined how the treatment effect from the field experiment differs according the variables that were shown to albeit mildly determine selection. We find some evidence of significant differences in the treatment effect although these results are mixed and inconsistent.

Overall we find that the schools that are part of the experimental sample do not show strong systematic differences along observable socio-economic characteristics, but they are those who have possibly experimented before and are more likely to be located in areas where there is a problem to solve.

While this study is merely a show case of the type of selection that could take place in field experimental research, we believe it brings a general positive message: Researchers could in principle afford to draw broader conclusions from their studies if more effort was spent on documenting the nature of selection into the experiment. Presenting summary statistics like those reported in Table 1 should be relatively straightforward in most field experimental studies. If a sufficient number of partners has been selected, it may even be possible to study how the treatment effects vary with observables, in the spirit of Table 8. Although both exercises are imperfect as they rely on observables only, they would in opinion largely improve the current situation, where researchers are forced to be extremely conservative in drawing conclusions and extrapolate to the general population.

References

Allcott, H., (2015). "Site Selection Bias in Program Evaluation", *Quarterly Journal of Economics* Vol. 130, No. 3 (August), pp. 1117-1165.

Alatas, V., Banerjee, A., Hanna, R., Olken B.A., and Tobias, J. (2012). "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review*, 102(4): 1206-40.

Altonji, J., Elder, T., and Taber, C. (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, Vol. 113, pp. 151–184.

Ambler, K., Aycinena, D., and Yang, D. (2015). "Channeling Remittances to Education: A Field Experiment among Migrants from El Salvador." *American Economic Journal: Applied Economics*, 7(2): 207-32.

Attanasio, O., Augsburg, B., Haas, R.D, Fitzsimons, E., and Harmgart, H., (2015). "The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia." *American Economic Journal: Applied Economics*, 7(1): 90-122.

Avvisati, F., Gurgand, M., Guyon, N., and Maurin, E., (2014). "Getting Parents Involved: A Field Experiment in Deprived Schools," *Review of Economic Studies*, vol. 81(1), pages 57-83.

Bandiera, O., Barankay, I., Rasul, I., (2007). "Incentives for Managers and Inequality Among Workers: Evidence From a Firm-Level Experiment", *Quarterly Journal of Economics*, Vol. 122(2), pp. 729-773.

Belot, M., and James, J., (2014). "A new perspective on the issue of selection bias into randomized controlled field experiments", *Economics Letters*, Vol. 124(3), pp. 326-328.

Belot, M., James, J., and Nolen, P., (2015). "Incentives and Children's Dietary Choices: A Field Experiment in Primary Schools", IZA Discussion paper no. 9424

Bettinger, E.P., Long, B.T., Oreopoulos, P., and Sanbonmatsu, L., (2012) "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment" *The Quarterly Journal of Economics*, 127 (3): 1205-124

Blimpo, M. P., (2014). "Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin." *American Economic Journal: Applied Economics*, 6(4): 90-109.

Burtless, G. (1995). "The Case for Randomized Field Trials in Economic and Policy Research", *Journal of Economic Perspectives*, Vol. 9(2), pp. 63-84.

Bursztyn, L., and Jensen, R., (2015). "How Does Peer Pressure Affect Educational Investments?" *The Quarterly Journal of Economics*, 130 (3): 1329-1367

Cai, H., Chen, Y., and Fang, H., (2009). "Observational Learning: Evidence from a Randomized Natural Field Experiment." *American Economic Review*, 99(3): 864-82.

Card, D., DellaVigna, S., and Malmendier, U. (2011). "The Role of Theory in Field Experiments", *Journal of Economic Perspectives*, Vol. 25(3), pp. 39-62.

Chetty, R., Looney, A., and Kroft, K., (2009). "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145-77.

Cleave B., N. Nikiforakis and Slonim, R., (2013), "Is there selection bias in laboratory experiments? The case of social and risk preferences", *Experimental Economics*, Vol. 16, Issue 3, pp. 372-382

de Mel, S., McKenzie, D., and Woodruff, C., (2013). "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics*, 5(2): 122-50.

Duflo, E., and Kremer, M. (2005). "Use of Randomization in the Evaluation of Development Effectiveness", in George Pitman, Osvaldo Feinstein and Gregory Ingram (eds.) Evaluating Development Effectiveness. New Brunswick, NJ: Transaction Publishers.

Duflo, E., Dupas, P., & M. Kremer (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya", *American Economic Review*, Vol. 101(5), pp. 1739-1774

Duflo, E., Greenstone, M., and Ryan, N., (2013). "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India," *The Quarterly Journal of Economics*, Oxford University Press, vol. 128(4), pages 1499-1545.

Duflo, E., Gale, W., Liebman, J., Orszag, P., and Saez, E., (2006). "Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block," *The Quarterly Journal of Economics*, vol. 121(4), pages 1311-1346, November.

Dupas, P. (2011). "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya", *American Economic Journal: Applied Economics,* Vol. 3(1), pp. 1-34.

Dupas, P., and Robinson, J., (2013). "Why Don't the Poor Save More? Evidence from Health Savings Experiments." *American Economic Review*, 103(4): 1138-71.

Falk, A. (2007). "Gift Exchange in the Field", *Econometrica*, Vol. 75(5), pp. 1501-1511.

Fehr, E. and Goette, L. (2007). "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment", *American Economic Review,* Vol. 97(1), pp. 298-317.

Field, E., Pande, R., Papp, J., and Rigol, N.,. (2013). "Does the Classic Microfinance Model Discourage Entrepreneurship among the Poor? Experimental Evidence from India." *American Economic Review*, 103(6): 2196-2226.

Frijters, Paul, Kong, T. and Liu, E. (2015). "Who is coming to the artefactual field experiment? Participation bias among Chinese rural migrants." *Journal of Economic Behavior and Organization*, 114, pp. 62-74.

Fryer Jnr, R.G., (2011). "Financial Incentives and Student Achievement: Evidence from Randomized Trials", *Quarterly Journal of Economics*, Vol. 126(4), pp. 1755-1798.

Fryer Jnr., R.G., (2014). "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments" *The Quarterly Journal of Economics*, 129 (3): 1355-1407

Gautier, P. and van der Klaauw, B., (2012). "Selection in a field experiment with voluntary participation", *Journal of Applied Econometrics*, Vol. 27, pp. 63-84.

Gine, X., Goldberg, J., and Yang, D., (2012). "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi." *American Economic Review*, 102(6): 2923-54.

Hastings, J., and Weinstein, J. (2008). "School Choice, and Academic Achievement: Evidence from Two Experiments", *Quarterly Journal of Economics*, Vol. 123(4), pp. 1373-1414

Harrison, G. W., and List, J.A., (2004). "Field experiments". *Journal of Economic Literature* Vol. 42, pp. 1009–105

Heckman, J.J., and Vytlacil, E.J. (2001). "Policy-Relevant Treatment Effects", *The American Economic Review* Vol. 91(2), pp. 107-111

Heckman, J.J., and Vytlacil, E.J. (2006). "Econometric Evaluation of Social Programs". *Handbook of Econometrics*, Volume 6, J. Heckman and E. Leamer, eds., Amsterdam: North Holland.

Heckman, J.J., Lalonde, R., and Smith, J. (1999). "The Economics and Econometrics of Active Labor Market Programs", *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and D. Card, eds., Amsterdam: Elsevier

Holt, C. A. (2005). *Markets, Games and Strategic Behavior: Recipes for Interactive Learning.* Addison-Wesley, Boston

Hotz, J. V., Imbens, G. W., and Mortimer, J. H., (2005). "Predicting the efficacy of future training programs using past experiences at other locations," *Journal of Econometrics*, vol. 125(1-2), pp. 241-270.

Jones, D., (2010). "Information, Preferences, and Public Benefit Participation: Experimental

Evidence from the Advance EITC and 401(k) Savings." *American Economic Journal: Applied Economics*, 2(2): 147-63.

Dean, K., and List, J.A., (2007). "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment." *American Economic Review*, 97(5): 1774-1793.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011), "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica*, 79: 651692.

List, J. (2011). "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off." *Journal of Economic Perspectives*, Vol. 25(3): pp. 3–16.

Ludwig, J., Kling, J.R., and Mullainathan, S.,. 2011. "Mechanism Experiments and Policy Evaluations". *Journal of Economic Perspectives*, Vol. 25(3), pp. 17–38.

Nagin, D.S., Rebitzer, J.B., Sanders, S., and Taylor, L.J., (2002). "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *American Economic Review*, 92(4): 850-873.

Oster, E., and Thornton, R., (2011). "Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation", *American Economic Journal: Applied Economics* 3(1): pp. 91-100

Pradhan, M., Suryadarma, D, Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., and Artha, R.P., (2014). "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia." *American Economic Journal: Applied Economics*, 6(2): 105-26.

Royer, H., Stehr, M., and Sydnor, J., (2015). "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics*, 7(3): 51-84.

Shearer, B., (2004). "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment" *The Review of Economic Studies* Vol. 71, No. 2 , pp. 513-534

Slonim, R., Wang, C., Garbarino, E., and Merrett, D. (2013). "Opting-in: Participation bias in economic experiments." *Journal of Economic Behavior and Organization*, 90, 43-70.

Wisdom, J., Downs, J.S., and Loewenstein, G., (2010). "Promoting Healthy Choices: Information versus Convenience", *American Economic Journal: Applied Economics*, 2(2), pp. 164-78.

Figure 1: Decision tree of the local authorities

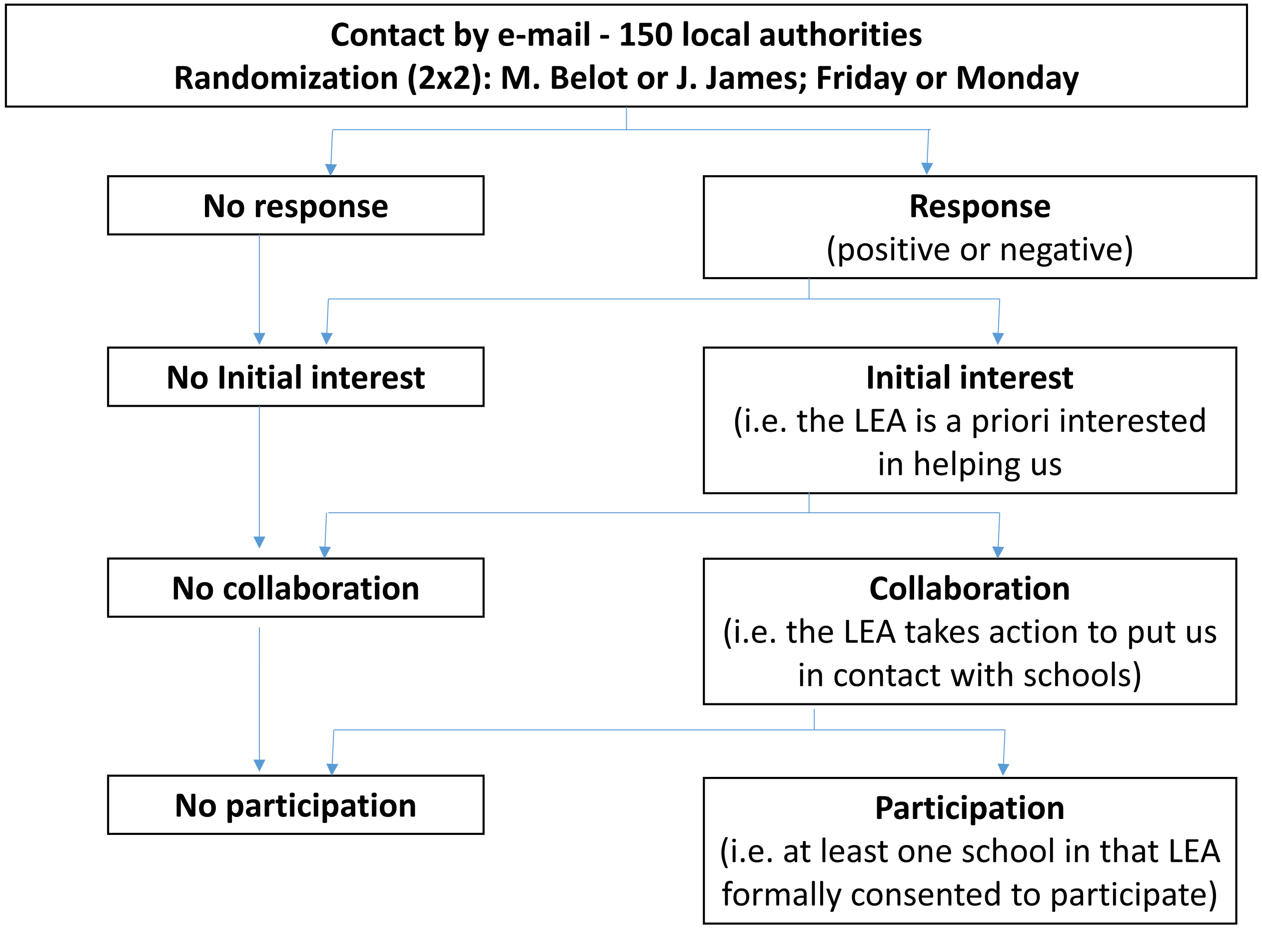

Table 1: Descriptive Statistics

	All		Responded		Did not Respond			Participate		Did not Participate		
	mean	s.d.	mean	s.d.	mean	s.d.	p-value	mean	s.d.	mean	s.d.	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Contacted on Friday	0.49	0.50	0.59	0.50	0.41	0.50	0.04	0.75	0.45	0.47	0.50	0.06
Contacted by J James	0.51	0.50	0.56	0.50	0.48	0.50	0.34	0.58	0.51	0.50	0.50	0.60
Household Income/100	6.72	1.45	6.72	1.42	6.73	1.48	0.95	6.10	0.81	6.78	1.48	0.12
% FSM	0.18	0.09	0.18	0.10	0.17	0.08	0.26	0.15	0.07	0.18	0.09	0.33
Number of Schools/100	1.45	1.23	1.64	1.42	1.30	1.06	0.10	2.27	2.06	1.37	1.11	0.02
% Eat 5 Fruit & Veg a day	25.9	4.9	25.4	4.8	26.2	5.0	0.34	24.5	4.4	26.0	4.9	0.31
% Overweight & Obese reception	23.0	2.7	23.2	2.8	22.9	2.6	0.47	23.6	3.1	23.0	2.6	0.44
% Obese reception	9.84	1.76	10.13	1.90	9.62	1.62	0.08	9.94	1.67	9.83	1.77	0.83
% Overweight & Obese yr 6	33.2	3.3	33.6	3.5	32.9	3.1	0.18	33.1	3.3	33.2	3.3	0.93
% Obese yr 6	18.8	2.9	19.2	3.2	18.5	2.7	0.16	18.5	2.5	18.9	3.0	0.65
% Smoking	24.7	4.6	25.0	4.6	24.5	4.6	0.58	25.6	5.1	24.6	4.5	0.49
% Binge Drinking	18.2	4.6	18.1	4.4	18.3	4.7	0.82	20.2	3.6	18.0	4.6	0.12
Key stage 1: Avg point score	0.15	0.01	0.15	0.01	0.15	0.01	0.54	0.15	0.00	0.15	0.01	0.61
Key stage 2: Avg point score	0.28	0.00	0.28	0.00	0.28	0.00	0.76	0.28	0.00	0.28	0.00	0.91
Per pupil spending 2010/11	4509	583	4582	674	4452	498	0.19	4307	220	4527	602	0.21
% change in per pupil spending 2010/11	0.04	0.00	0.04	0.00	0.04	0.00	0.62	0.04	0.00	0.04	0.00	0.51
% LA spending change 2010/11	-0.05	0.03	-0.05	0.03	-0.05	0.03	0.56	-0.05	0.03	-0.05	0.03	1.00
Female CEO of the council	0.26	0.44	0.23	0.42	0.28	0.45	0.44	0.25	0.45	0.26	0.44	0.94
Female Director of Children Services	0.51	0.50	0.56	0.50	0.48	0.50	0.34	0.67	0.49	0.50	0.50	0.26
Female Leader of Healthy Schools	0.82	0.38	0.83	0.38	0.82	0.39	0.90	0.75	0.45	0.83	0.38	0.51
% of Labour Councillors	0.33	0.26	0.34	0.26	0.32	0.26	0.65	0.32	0.25	0.33	0.26	0.83
% of Conservative Councillors	0.42	0.26	0.41	0.25	0.43	0.27	0.76	0.43	0.25	0.42	0.26	0.87
Labour controlled council	0.29	0.46	0.30	0.46	0.28	0.45	0.78	0.25	0.45	0.29	0.46	0.75
Conservative controlled council	0.42	0.50	0.41	0.50	0.43	0.50	0.87	0.42	0.51	0.42	0.50	0.98
Ofsted Score	2.29	0.20	2.30	0.19	2.28	0.21	0.47	2.34	0.25	2.28	0.19	0.34
Ofsted Health Score	1.70	0.16	1.71	0.16	1.69	0.16	0.49	1.68	0.16	1.70	0.16	0.71
Catering pp/100	0.82	0.59	0.85	0.63	0.80	0.57	0.68	0.80	0.63	0.82	0.59	0.88
Energy costs pp/100	0.66	0.12	0.67	0.12	0.65	0.12	0.43	0.69	0.11	0.66	0.12	0.31
Total school Income pp/1000	4.34	0.50	4.43	0.53	4.28	0.47	0.07	4.29	0.26	4.35	0.52	0.73
Teaching costs pp/1000	2.14	0.18	2.17	0.18	2.11	0.18	0.07	2.10	0.11	2.14	0.18	0.51

Notes: p-values in column 7 refer to a two-sided t-test between local authorities that responded against those that did not respond, and column 12 shows the p-value that tests the difference between those local authorities that had schools which participated in the experiment versus those that did not. Local authorities were randomly contacted on two days on Friday 2nd July and Monday 5th July and by J. James or M. Belot. Income is the average weekly total household income (£) divided by 100, FSM is the percentage of children who are eligible for free school meals. % Eat 5 Fruit & Veg a day is the proportion of adults defined to be consumers of 5 or more fruit and vegetables if they had reported that they had consumed 5 or more portions of fruit and vegetables on the previous day. Binge drinking is the proportion of adult binge drinkers defined if they reported that in the last week they had drunk 8 or more units of alcohol if they were a man, or 6 or more units of alcohol if they were a woman, on any one day or more. Smoking is the proportion of individuals in a local authority who reported that they were a 'current cigarette smoker' in the Health Survey for England. Overweight and Obese reception is the percentage of pupils in the local authority who were overweight or obese when they entered primary school aged 4 or 5. Year 6 is the final year of school when the pupils are aged 10 or 11. The average point score (APS) of the key stage 1 test and key stage 2 point score are for tests taken in primary school. The points are awarded per subject per pupil along the following lines: working below the level of the test or not awarded 15, level 2 receives 15 points, level 3 gets 21, level 4 gets 27 and 33 points is allocated for level 5. The APS is then calculated using the following: (Total points for English + Total points for maths + Total points for science) / (Total number of eligible pupils for each subject). This is then rescaled by dividing by 100. Per pupil spending in 2010/11, the yearly increase in per pupil spending, and the overall change in the spending of the local authority. Labour Party and Conservative Party councillors on the council defined at the most recent election since July 2010. Ofsted is (the government school inspector) average score of the schools in the local authority. Schools are inspected and judged on the following question: "How effective, efficient and inclusive is the provision of education, integrated care and any extended services in meeting the needs of learners?" With ratings given of: 1. Outstanding 2. Good 3. Satisfactory 4. Inadequate. Ofsted Health Score is based on the following question: "Learners are encouraged and enabled to eat and drink healthily" using the same 1 to 4 scale. Average catering costs (including staff costs), energy, teaching and total school income are per pupil averages at the local authority level and are rescaled as indicated.

Table 2 Main Determinants of response to initial e-mail, interest and response

	Panel A: Response					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.142	0.147*	0.141	0.140	0.124	0.123
	(0.087)	(0.087)	(0.087)	(0.087)	(0.086)	(0.087)
Contacted by J James	0.076	0.073	0.075	0.072	0.085	0.085
	(0.085)	(0.085)	(0.085)	(0.085)	(0.084)	(0.084)
Income/100	0.084*	0.082	0.084*	0.085*	0.016	0.007
	(0.050)	(0.050)	(0.051)	(0.050)	(0.038)	(0.038)
% FSM	0.741	0.129	0.660	0.344	0.759	0.843
	(0.570)	(0.727)	(0.782)	(0.808)	(0.591)	(0.525)
Number of schools/100	0.080**	0.080**	0.080**	0.081**	0.071*	0.070*
	(0.039)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)
% Eat 5 Fruit & Veg a day	-0.028*	-0.026*	-0.028*	-0.027*		
	(0.015)	(0.015)	(0.016)	(0.015)		
% Overweight & Obese Reception	-0.000					
	(0.019)					
% Obese Reception		0.044				
		(0.036)				
% Overweight & Obese Year 6			0.003			
			(0.021)			
% Obese Year 6				0.016		
				(0.025)		
% Smokers					0.004	
					(0.014)	
% Binge Drinkers						-0.001
						(0.012)
Observations	145	145	145	145	145	145
R squared	0.0656	0.0731	0.0657	0.0676	0.0482	0.0479
	Panel B: Interest					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.066	0.070	0.043	0.044	0.046	0.052
	(0.080)	(0.080)	(0.081)	(0.081)	(0.080)	(0.082)
Contacted by J James	0.041	0.030	0.034	0.027	0.056	0.053
	(0.077)	(0.077)	(0.077)	(0.077)	(0.078)	(0.078)
Income/100	0.140***	0.134***	0.133***	0.143***	0.057*	0.035
	(0.047)	(0.047)	(0.048)	(0.048)	(0.034)	(0.034)
% FSM	0.568	-0.057	-0.122	-0.380	0.566	0.813*
	(0.479)	(0.610)	(0.655)	(0.717)	(0.497)	(0.458)
Number of schools/100	0.087***	0.089***	0.090***	0.095***	0.074**	0.069**
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
% Eat 5 Fruit & Veg a day	-0.036**	-0.035**	-0.033**	-0.035**		
	(0.015)	(0.015)	(0.015)	(0.015)		
% Overweight & Obese Reception	0.017					
	(0.018)					
% Obese Reception		0.061*				
		(0.033)				
% Overweight & Obese Year 6			0.035*			
			(0.020)			
% Obese Year 6				0.051**		
				(0.026)		
% Smokers					0.015	
					(0.012)	
% Binge Drinkers						0.003
						(0.011)
Observations	128	128	128	128	128	128
R squared	0.113	0.131	0.128	0.133	0.0657	0.0565

Notes: Panel A presents the determinants of any response to the initial email that was sent to local authorities. Panel B shows the results for the local authority expressing an interest in the project. The coefficients are the marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1 Local authorities were randomly contacted on two days on Friday 2nd July and Monday 5th July and by J. James or M. Belot. Income is the average weekly total household income (£) divided by 100, FSM is the percentage of children who are eligible for free school meals. % Eat 5 Fruit & Veg a day is the proportion of adults defined to be consumers of 5 or more fruit and vegetables if they had reported that they had consumed 5 or more portions of fruit and vegetables on the previous day. Binge drinkers is the proportion of adult binge drinkers defined if they reported that in the last week they had drunk 8 or more units of alcohol if they were a man, or 6 or more units of alcohol if they were a woman, on any one day or more. Smokers is the proportion of individuals in a local authority who reported that they were a 'current cigarette smoker' in the Health Survey for England. Overweight and Obese represent reception is the percentage of pupils in the local authority who were overweight or obese when they entered primary school aged 4 or 5. Year 6 is the final year of school when the pupils are aged 10 or 11.

Table 3 Main Determinants of response to initial e-mail, collaboration and participation

	Panel A: Collaboration					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.004	0.007	-0.007	-0.006	-0.010	0.000
	(0.067)	(0.066)	(0.067)	(0.066)	(0.068)	(0.068)
Contacted by J James	0.014	0.011	0.010	0.008	0.022	0.021
	(0.064)	(0.064)	(0.064)	(0.064)	(0.065)	(0.065)
Income/100	0.086**	0.083**	0.081**	0.086**	0.030	0.030
	(0.040)	(0.039)	(0.040)	(0.040)	(0.029)	(0.029)
% FSM	0.425	0.055	0.092	-0.141	0.437	0.580
	(0.408)	(0.517)	(0.563)	(0.595)	(0.425)	(0.382)
Number of schools/100	0.071***	0.072***	0.072***	0.075***	0.063**	0.062**
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
% Eat 5 Fruit & Veg a day	-0.024*	-0.024*	-0.022*	-0.023*		
	(0.012)	(0.012)	(0.012)	(0.012)		
% Overweight & Obese Reception	0.009					
	(0.015)					
% Obese Reception		0.035				
		(0.026)				
% Overweight & Obese Year 6			0.017			
			(0.016)			
% Obese Year 6				0.028		
				(0.019)		
% Smokers					0.008	
					(0.010)	
% Binge Drinkers						0.007
						(0.009)
Observations	145	145	145	145	145	145
R squared	0.0748	0.0840	0.0792	0.0864	0.0460	0.0458

	Panel B: Participation					
	(1)	(2)	(3)	(4)	(5)	(6)
Contacted on Friday	0.062	0.064*	0.058	0.061	0.060	0.064
	(0.040)	(0.038)	(0.040)	(0.040)	(0.040)	(0.042)
Contacted by J James	0.014	0.012	0.011	0.012	0.013	0.012
	(0.032)	(0.028)	(0.033)	(0.033)	(0.033)	(0.033)
Income/100	-0.029	-0.030	-0.033	-0.031	-0.029	-0.019
	(0.022)	(0.020)	(0.023)	(0.023)	(0.019)	(0.021)
% FSM	-0.402	-0.583**	-0.520	-0.435	-0.448	-0.396
	(0.282)	(0.288)	(0.354)	(0.386)	(0.322)	(0.279)
Number of schools/100	0.015	0.015	0.015	0.015	0.015	0.015
	(0.013)	(0.012)	(0.013)	(0.014)	(0.013)	(0.013)
% Eat 5 Fruit & Veg a day	-0.000	-0.001	-0.000	-0.002		
	(0.006)	(0.005)	(0.007)	(0.007)		
% Overweight & Obese Reception	0.009					
	(0.008)					
% Obese Reception		0.024*				
		(0.013)				
% Overweight & Obese Year 6			0.007			
			(0.009)			
% Obese Year 6				0.003		
				(0.011)		
% Smokers					0.004	
					(0.005)	
% Binge Drinkers						0.007
						(0.006)
Observations	145	145	145	145	145	145
R squared	0.159	0.185	0.150	0.144	0.148	0.160

Notes: Panel A examines whether the local authority collaborated by providing the names of schools. Panel B examines whether a local authority had a school take part in the experiment. The coefficients are the marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1 Local authorities were randomly contacted on two days on Friday 2nd July and Monday 5th July and by J. James or M. Belot. Income is the average weekly total household income, FSM is the percentage of children who are eligible for free school meals. Fruit & Veg is the proportion of adults defined to be consumers of 5 or more fruit and vegetables if they had reported that they had consumed 5 or more portions of fruit and vegetables on the previous day. Binge drinking is the proportion of adult binge drinkers defined if they reported that in the last week they had drunk 8 or more units of alcohol if they were a man, or 6 or more units of alcohol if they were a woman, on any one day or more. Smoking is the proportion of individuals in a local authority who reported that they were a 'current cigarette smoker' in the Health Survey for England. Overweight and Obese represent reception is the percentage of pupils in the local authority who were overweight or obese when they entered primary school aged 4 or 5. Year 6 is the final year of school when the pupils are aged 10 or 11.

Table 4 Role of Pre-Trends in Obesity and Education

	Panel A: Obesity Trends							
	Response		Interest		Collaboration		Participation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted by J James	0.090	0.091	0.059	0.060	0.020	0.021	0.014	0.015
	(0.084)	(0.084)	(0.076)	(0.076)	(0.064)	(0.064)	(0.035)	(0.035)
Income/100	0.076	0.073	0.129***	0.126***	0.081**	0.079**	-0.034	-0.032
	(0.050)	(0.050)	(0.047)	(0.048)	(0.039)	(0.040)	(0.024)	(0.024)
% FSM	0.823	0.636	0.739	0.561	0.514	0.370	-0.269	-0.180
	(0.588)	(0.737)	(0.511)	(0.644)	(0.440)	(0.538)	(0.333)	(0.381)
Number of schools/100	0.093**	0.094**	0.091***	0.093***	0.071***	0.072***	0.022	0.022
	(0.038)	(0.038)	(0.032)	(0.032)	(0.027)	(0.027)	(0.014)	(0.014)
% Eat 5 Fruit and Veg a day	-0.025	-0.025	-0.037**	-0.037**	-0.024**	-0.025**	-0.002	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)	(0.012)	(0.012)	(0.007)	(0.007)
Obesity 2008-Obesity 2006 (Reception)	-0.005		-0.005		0.000		-0.000	
	(0.022)		(0.020)		(0.017)		(0.009)	
Obesity 2008-Obesity 2006 (y6)		-0.009		-0.009		-0.005		0.003
		(0.019)		(0.018)		(0.015)		(0.008)
Observations	146	146	129	129	146	146	146	146
R squared	0.0508	0.0517	0.0991	0.100	0.0702	0.0711	0.111	0.113
	Panel B: Education Trends							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted by J James	0.092	0.089	0.034	0.030	0.011	0.008	0.017	0.016
	(0.085)	(0.084)	(0.077)	(0.077)	(0.063)	(0.063)	(0.032)	(0.032)
Income/100	0.078	0.074	0.137***	0.139***	0.081**	0.085**	-0.032	-0.034
	(0.050)	(0.050)	(0.048)	(0.048)	(0.040)	(0.039)	(0.022)	(0.022)
% FSM	0.454	0.313	-0.405	-0.252	-0.250	-0.108	-0.256	-0.309
	(0.843)	(0.855)	(0.733)	(0.755)	(0.613)	(0.612)	(0.367)	(0.384)
Number of schools/100	0.094**	0.092**	0.098***	0.097***	0.073***	0.073***	0.026*	0.024*
	(0.038)	(0.038)	(0.032)	(0.031)	(0.026)	(0.026)	(0.014)	(0.013)
% Obese Reception	0.040	0.040	0.030	0.032	0.020	0.019	0.025	0.026
	(0.043)	(0.043)	(0.040)	(0.039)	(0.032)	(0.032)	(0.017)	(0.017)
% Obese Year 6	-0.001	-0.000	0.037	0.036	0.019	0.021	-0.006	-0.007
	(0.030)	(0.030)	(0.031)	(0.031)	(0.023)	(0.024)	(0.012)	(0.012)
KS1 Score 2009-KS1 Score 2007	-0.155		-0.052		0.055		-0.085	
	(0.226)		(0.199)		(0.165)		(0.090)	
% Eat 5 Fruit and Veg a day	-0.022	-0.022	-0.032**	-0.033**	-0.022*	-0.022*	-0.001	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)	(0.012)	(0.012)	(0.006)	(0.006)
KS2 Score 2009-KS2 Score 2007		3.125		-20.65		-15.45		-3.82
		(24.33)		(22.77)		(18.40)		(10.11)
Observations	146	146	129	129	146	146	146	146
R squared	0.0592	0.0569	0.131	0.136	0.0870	0.0912	0.155	0.145

Notes: Panel A examines whether the change in obesity rates determine whether a local authority responds (columns 1 and 2), expresses interest (columns 3 and 4), collaborates (columns 5 and 6), or participates (columns 7 and 8). In the odd columns we focus on those just entering primary school, and in even columns we examine those in the final year of primary school. Panel B examines changes in education performance. In the odd columns we focus on key stage 1 exams (taken aged 6), and in even columns we examine key stage 2 (taken aged 10). The coefficients are the marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 5 Splines of key variables

	Response		Interest		Collaboration		Participation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contacted on Friday	0.125	0.120	0.039	0.044	-0.009	-0.008	0.017	0.011
	(0.087)	(0.088)	(0.080)	(0.080)	(0.065)	(0.065)	(0.020)	(0.015)
Contacted by J James	0.099	0.116	0.069	0.074	0.023	0.034	-0.001	0.003
	(0.085)	(0.086)	(0.077)	(0.078)	(0.063)	(0.063)	(0.010)	(0.007)
% FSM	1.137	0.210	0.476	0.009	0.305	-0.021	-0.076	-0.120
	(0.728)	(0.651)	(0.633)	(0.554)	(0.513)	(0.447)	(0.108)	(0.129)
Number of schools/100	0.072*	0.084**	0.083**	0.086**	0.066**	0.071**	0.000	0.001
	(0.040)	(0.040)	(0.034)	(0.034)	(0.028)	(0.028)	(0.004)	(0.003)
Income/100	-0.003	0.018	0.082	0.077	0.053	0.055	-0.049	-0.035
	(0.064)	(0.065)	(0.056)	(0.057)	(0.047)	(0.047)	(0.037)	(0.033)
Income Mid	0.180	0.133	0.162	0.104	0.107	0.067	0.049	0.030
	(0.128)	(0.131)	(0.129)	(0.130)	(0.108)	(0.106)	(0.049)	(0.036)
Income Upper	0.228	0.106	0.205	0.110	0.170	0.095	0.616**	0.508
	(0.216)	(0.227)	(0.215)	(0.216)	(0.189)	(0.182)	(0.283)	(0.316)
Fruit & Veg Mid	-0.118	-0.084	-0.099	-0.069	-0.086	-0.070	0.018	0.023
	(0.120)	(0.122)	(0.100)	(0.106)	(0.079)	(0.081)	(0.026)	(0.029)
Fruit & Veg Upper	-0.210	-0.174	-0.323***	-0.287**	-0.266***	-0.246***	-0.003	0.001
	(0.148)	(0.149)	(0.110)	(0.116)	(0.092)	(0.094)	(0.019)	(0.015)
Obesity Mid (y6)	-0.010		0.233*		0.107		-0.007	
	(0.117)		(0.125)		(0.102)		(0.011)	
Obesity Upper (y6)	-0.084		0.158		0.093		-0.016	
	(0.152)		(0.168)		(0.135)		(0.021)	
Obese Mid (Reception)		0.023		0.026		0.005		-0.008
		(0.116)		(0.112)		(0.088)		(0.011)
Obese Upper (Reception)		0.236*		0.286*		0.204		0.009
		(0.142)		(0.157)		(0.135)		(0.018)
Observations	147	147	130	130	147	147	147	147
R squared	0.0678	0.0831	0.144	0.152	0.109	0.129	0.289	0.318

Notes: All coefficients presented as marginal effects from a probit regression, standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Mid refers to above the 33rd and below (and equal to) the 66th centiles, upper refers to above the 66th percentile, the r eference group is below (or equal to) the 33rd pecentile. Overweight and Obese reception is the percentage of pupils in the local authority who were overweight or obese when they entered primary school aged 4 or 5. Year 6 is the final year of school when the pupils are aged 10 or 11.

Table 6 Selected Schools

	(1)	(2)	(3)	(4)	(5)
		Selected schools		Experiment sample	
Ofsted Health Score 1	0.0147*	0.0169**	0.0169**	0.0107	0.0118*
	(0.00781)	(0.00837)	(0.00835)	(0.00655)	(0.00667)
Ofsted Health Score 3 or 4	0.0262	0.0268	0.0268	0.0182	0.0193
	(0.0172)	(0.0177)	(0.0176)	(0.0145)	(0.0141)
Ofsted Score 1	-0.0109	-0.0297**	-0.0298***	-0.00386	-0.0230**
	(0.0101)	(0.0116)	(0.0114)	(0.00849)	(0.00926)
Ofsted Score 3 or 4	-0.00744	-0.00770	-0.00769	-0.00110	0.000690
	(0.00763)	(0.00818)	(0.00799)	(0.00640)	(0.00652)
FSM Medium	0.0277**	0.0252*	0.0252*	0.0212*	0.0167
	(0.0139)	(0.0146)	(0.0145)	(0.0116)	(0.0116)
FSM Low	0.00657	0.00514	0.00512	0.00545	0.00132
	(0.0134)	(0.0149)	(0.0144)	(0.0112)	(0.0119)
Energy Costs per pupil/100	0.0156	0.0151	0.0151	0.000138	-0.00479
	(0.0114)	(0.0127)	(0.0127)	(0.00959)	(0.0102)
Total School Income per pupil/1000	-0.00305	-0.00649	-0.00648	0.00135	-0.000778
	(0.00719)	(0.00846)	(0.00821)	(0.00603)	(0.00674)
Catering (inc staff) costs per pupil/100	-0.00379	-0.00560	-0.00560	6.93e-05	-0.000444
	(0.00506)	(0.00595)	(0.00595)	(0.00424)	(0.00474)
Total pupils/100	0.00111	0.00243	0.00244	-0.00263	-0.00104
	(0.00364)	(0.00397)	(0.00390)	(0.00305)	(0.00316)
Teaching Costs per pupil/1000	-0.00500	0.00859	0.00858	-0.0163	-0.00894
	(0.0135)	(0.0156)	(0.0156)	(0.0113)	(0.0124)
Avg Eng/Math Score 09/10		-0.00144			0.0783
		(0.277)			(0.221)
Constant	0.0577	0.0490	0.0486	0.0815***	0.0609
	(0.0363)	(0.0937)	(0.0428)	(0.0304)	(0.0747)
Observations	2,087	1,778	1,778	2,087	1,778
R-squared	0.045	0.045	0.045	0.031	0.032
Mean of dep. var	0.022	0.021	0.021	0.015	0.013

Notes: The sample is all schools in the participating local authorities. Ofsted is (the government school inspector) average score of the schools in the local authority. Schools are inspected and judged on the following question: "How effective, efficient and inclusive is the provision of education, integrated care and any extended services in meeting the needs of learners?" With ratings given of: 1. Outstanding 2. Good 3. Satisfactory 4. Inadequate. Ofsted Health Score is based on the following question: "Learners are encouraged and enabled to eat and drink healthily" using the same 1 to 4 scale. Average catering costs (including staff costs), energy, teaching and total school income are per pupil averages at the local authority level and are rescaled as indicated. FSM Band - The three broad bands used to group pupils eligible for FSM are: Low: less than 20%, Medium: 20.01-35% and High: greater than 35% (omitted catergory). Columns (1)-(3) present estimates using whether a school was selected by the LEA. Column (3) excludes "Avg Eng/Math Score" but uses the same sample in column (2). Column (4) and (5) use whether a school started and completed the experimental intervention. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 7 Comparison of participating schools from the pool of selected schools

	Experiment	Dropped	p-value of difference
% Girls	0.48	0.49	0.802
Number of pupils	207	279	0.322
Total School Income per pupil/1000	4.17	4.16	0.641
Catering costs per pupil/100	0.96	0.73	0.303
% English and Maths above level 4 KS2	0.76	0.76	0.949
Average point score Maths and English	0.28	0.28	0.396
Ofsted Score	2.09	2.29	0.521
Ofsted health Score	1.53	1.43	0.604
FSM Medium	0.29	0.29	0.975
FSM Low	0.65	0.64	0.988
Teaching Costs per pupil/1000	2.05	2.17	0.246
Energy costs per pupil/100	0.64	0.87	0.961
Competition treatment	0.29	0.43	0.368
Individual treatment	0.32	0.36	0.822
Control	0.39	0.21	0.260
Schools	31	15	

Notes: Columns 1 and 2 show the mean values at the school level. Column 3 is the p-value of (Prob>z, where z is the test statistic) from an Mann-Whitney U test. Ofsted is (the government school inspector) average score of the schools in the local authority. Schools are inspected and judged on the following question: "How effective, efficient and inclusive is the provision of education, integrated care and any extended services in meeting the needs of learners?" With ratings given of: 1. Outstanding 2. Good 3. Satisfactory 4. Inadequate. Ofsted Health Score is based on the following question: "Learners are encouraged and enabled to eat and drink healthily" using the same 1 to 4 scale. Average catering costs (including staff costs), energy, teaching and total school income are per pupil averages at the local authority level and are rescaled as indicated. FSM Band - The three broad bands used to group pupils eligible for FSM are: Low: less than 20%, Medium: 20.01-35% and High: greater than 35% (omitted catergory). Columns (1)-(3) present estimates using whether a school was selected by the LEA. Column (3) excludes "Avg Eng/Math Score" but uses the same sample in column (2). Column (4) and (5) use whether a school started and completed the experimental intervention.

Table 8 Treatment Effects and Selection Characteristics

	(1)	(2)	(3)	(4)
	Choice		Try	
	All	<100%	All	<100%
Comp x On	-0.074	-0.175***	-0.009	0.108*
	(0.048)	(0.036)	(0.058)	(0.056)
	[0.282]	[0.250]	[0.909]	[0.519]
Ind x On	-0.122	-0.214***	-0.087	-0.145**
	(0.106)	(0.067)	(0.109)	(0.053)
	[0.513]	[0.553]	[0.611]	[0.541]
Comp x On x Above Avg FV	0.084*	0.297***	0.262***	0.275***
	(0.042)	(0.024)	(0.041)	(0.029)
	[0.178]	[0.160]	[0.162]	[0.438]
Ind x On x Above Avg FV	0.092	0.075	0.138	0.109
	(0.116)	(0.146)	(0.100)	(0.090)
	[0.613]	[0.777]	[0.470]	[0.703]
On x Above Avg FV	-0.003	-0.024	-0.012	-0.038
	(0.024)	(0.023)	(0.035)	(0.024)
	[0.837]	[0.543]	[0.841]	[0.717]
Comp x On x (Health Score=1)	0.087*	0.221***	0.193***	0.194***
	(0.045)	(0.036)	(0.060)	(0.059)
	[0.238]	[0.00]	[0.27]	0.082
Ind x On x (Health Score=1)	0.029	-0.049	-0.033	-0.124**
	(0.104)	(0.067)	(0.109)	(0.053)
	[0.789]	[0.759]	[0.831]	[0.394]
On x (Health Score=1)	-0.105***	-0.097**	-0.048	0.011
	(0.025)	(0.035)	(0.050)	(0.038)
	[0.126]	[0.036]	[0.752]	[0.851]
Comp x On x Above Avg Income	-0.059	-0.229***	-0.306***	-0.366***
	(0.046)	(0.034)	(0.063)	(0.051)
	[0.294]	[0.000]	[0.000]	[0.020]
Ind x On x Above Avg Income	0.229***	0.352*	0.135	0.339***
	(0.066)	(0.170)	(0.093)	(0.103)
	[0.082]	[0.661]	[0.320]	[0.100]
On x Above Avg Income	-0.017	-0.036	0.058	0.011
	(0.031)	(0.032)	(0.052)	(0.034)
	[0.721]	[0.474]	[0.452]	[0.843]
Comp x On x Above Avg Number of Schools	0.081	0.278***	0.103*	0.075
	(0.052)	(0.023)	(0.058)	(0.052)
	[0.300]	[0.014]	[0.252]	[0.324]
Ind x On x Above Avg Number of Schools	-0.032	0.027	-0.012	-0.057
	(0.119)	(0.212)	(0.109)	(0.128)
	[0.877]	[0.947]	[0.897]	[0.791]
On x Above Avg Number of Schools	0.004	-0.080***	-0.001	-0.020
	(0.032)	(0.021)	(0.049)	(0.032)
	[0.921]	[0.074]	[1.000]	[0.621]
On	0.110***	0.297***	0.044	0.189***
	(0.025)	(0.035)	(0.050)	(0.038)
Constant	0.806***	0.496***	0.712***	0.414***
	(0.009)	(0.007)	(0.010)	(0.009)
Observations	11,759	4,464	11,218	4,364
R-squared	0.018	0.065	0.027	0.081
P-value of joint test of all triple interaction terms	0.00	0.00	0.00	0.00

Notes: This table presents the treatment effects from the field experiment briefly described in section 4.2 and more fully in Belot, James and Nolen (2013). Columns (1) and (2) use the dependent variable that indicates whether a fruit or vegetable was chosen, column (3) and (4) use the variable try. Columns (1) and (3) present estimates for all pupils, columns (2) and (4) restrict the sample to those who did not choose a fruit or vegetable each day in the baseline week before the incentives were in place. Robust standard errors clustered at school level are included in parentheses; * sig at 10%, ** sig at 5%, *** sig at 1%. Wild P-Values are shown in brackets and are estimated following Cameron, Gelbach, Miller (2008) using 1000 replications. Above Avg indicates where the fruit and vegetable consumption at the local authority level is equal to or above 26%. Health Score (= 1) indicates the ofsted score for adopting a healthy lifestyle was outstanding.